Topological navigation and Qualitative localization for indoor environment using Multisensory perception

P. Ranganathan†, J.B. Hayet‡, M. Devy‡, S. Hutchinson†, F. Lerasle‡

‡ LAAS-CNRS, Toulouse, France
E-mail : jbhayet, michel@laas.fr

† Beckman Institute at University of Illinois at Urbana-Champaign, Illinois, USA
E-mail : rangana, seth@uiuc.edu

Abstract. This article describes a navigation system for a mobile robot which must execute motions in a building; the robot is equipped with a belt of ultrasonic sensors and with a camera. The environment is represented by a topological model based on a Generalized Voronoi Graph and by a set of visual landmarks. Typically, the topological graph describes the free space in which the robot must navigate; a node is associated to an intersection between corridors, or to a crossing towards another topological area (an open space : rooms, hallways, ...); an edge corresponds to a corridor or to a path in an open space. Landmarks correspond to static, rectangular and planar objects (e.g. doors, windows, posters...) located on the walls. The landmarks are only located with respect to the topological graph : some of them are associated to nodes, other to edges. The paper is focused on the preliminary exploration task, i.e. the incremental construction of the topological model. The navigation task is based on this model : the robot self-localization is only expressed with respect to the graph.

1 Introduction

Navigation is a critical task for a mobile robot : it allows it to move and act autonomously in its environment. Because internal sensors on the robot (odometers) are not accurate enough or may give false measurements, a navigation system must be based on exteroceptive sensors like cameras, sonars or laser range finders. Many sensor-based navigation strategies have been proposed in the litterature. Only model-based strategies are considered here.

A first type of methods are based on explicit localization of the robot with respect to the environment [4, 3]; discriminant features or landmarks must be learnt during a preliminary exploration step; they are searched during the navigation step; the robot localization can be determined using a set of landmark matchings. The typical environment representation is a stochastic map for the landmark descriptions [15] and a grid-based or polygonal model for the free space. The robot position can be absolute (only one reference frame) or relative to an area (environment structuration in topologically independent areas : corridors, rooms...) [1]. A path for the robot is expressed by a curve in this free space; perceptual constraints can be taken into account by the path planner, so that landmarks are associated to each section of this path.

In methods proposed more recently [11, 16], the robot localization can be relative only to landmarks which are successively perceived by the robot; the continuity of a path is guaranteed by a graph -called a topological map- which expresses some relationships between landmarks; for example, landmark A is connected to landmark B only if B is visible from A, or if a sensor-based motion (wall following, visual servoing for example) can be executed to go from A to B.

This paper proposes a navigation system based on such a topological representation. For a service robot which must execute motions in an office environment (a corridors network that connects some open spaces like rooms or hallways), the topological map can be directly coupled to the environment structure, using the Generalized Voronoi Graph representation proposed by H.Choset [6, 12]; in such a graph, nodes are associated to transitions between areas (corridor crossings, area entrances, doors, ...); an edge typically corresponds to a path in a corridor or in an open space. For a corridor, the robot motion can be controlled using sonars to maintain the robot on the GVG; the robot localization is expressed according to the GVG (the robot is on this node or is moving on this edge).

Nevertheless, a self-localization problem may occur because this kind of environment is very ambiguous; using only sonars, the robot may recognize a false node due to an unexpected obstacle in a corridor close to an actual intersection (human presence) or to a topological modification (an open door). If the robot is equipped with several sensors -in our experiment, monocular vision and sonars-, it can take advantage of different topological representations (visual landmarks and GVG), in order to validate an hypothesis about a node recognition using some visual landmarks. Vision gives stable, reliable information from a large part of the environment, which may be helpful in comparison with ultrasonic sensors. It is used as the only sensor data to build topological graph in [5], for example.

Some authors have already presented results combining these two sensors in order to get a more robust representation. As an example, [8], the most similar work to ours, uses sonar by searching pre-determined forms of *gateways* and associates to these places some simple visual landmarks. However, the learning phase was not done autonomously, and the processing steps of sonar data made the algorithm usable in only orthogonal corridors. Our contribution is twofold :

- we make the robot *learn autonomously* the environment model, without preliminary guided route traversals and with extended environment structural configurations, although considering only corridors-based environment;
- we try to use a landmark intrinsic representation *independent from the viewpoint* and as stable as possible with respect to illumination, scale changes and small occlusions and combine it with some rough information about the viewing angle, so that exact calibration is not necessary.

The section 2 proposes an overview on the environment representation and the navigation system. The sections 3 and 4 present our strategy to build a hybrid topological map -landmark-based and GVG-; in the section 5, experimental results about the incremental construction of a model, are commented : we validated our method either in a very structured environment or in a more complex one. Finally in the section 6, discussions about this work and some future researches are considered.

2 Overview of our system

Topological maps, as proposed by Kuipers [9] represent the robot environment by graphs. The path is defined as a set of distinct points, which must be recognized by the robot using sensors data (mainly sonar data in the first results). These points provide the nodes of the map. Information about these point locations and characteristics are required to identify them; only this information need to be stored to traverse this map. If two nodes are directly connected, the connection is an arc in the map. The edges correspond to navigation operations such as wall following, visual servoing [14]... These navigation operations take the robot from one node to another. Such a representation has a lot of advantages :

- the node information may be meet points, landmarks detected by a vision system or any other distinguishing features of the environment that can be reliably recognized by a robot ;
- an advantage of topological maps is that the path between two nodes does not have to be traced exactly; it is sufficient if the robot can traverse a general path (not exactly defined) between two nodes. It is pivotal that the meet points are identified uniquely ;
- small storage capacities are required, since only information about the nodes are stored ;
- when the odometry is not used to localize the robot there is no need to maintain a global coordinate frame ;
- this method is suitable for exploring large-scale environments ;
- path planning from a topological map can be very fast and without complex computations.

However, topological maps may fail if two important properties are not guaranteed : first, the nodes have to be detected with certainty and accuracy, and secondly, the navigation operations must lead the robot from one node to another. In this paper, we focus on the first problem, a classical situation when two places look exactly alike (like corridors intersection in figure 11). Our aim is to link some other information in the graph so that failures in the node recognition could not occur. This visual information are currently added to the graph *nodes* only.

An example for a topological map is the Generalized Voronoi Graph (GVG), which has been popularized by Choset [6] with sonars; such a graph can also be built using a laser range finder [18]. The nodes are the so-called meet points and the arcs correspond to the Generalized Voronoi Diagram, i.e. the locus of all points

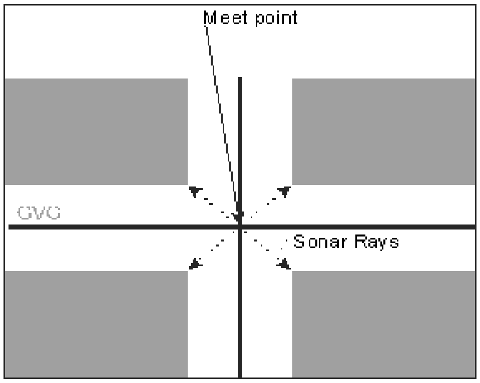


Fig. 1. *A meeting point*

that are equidistant to two object boundaries. At the intersection of two GVG edges there is a meet point, that is equidistant to at least, three points; only in special situations, it is possible that the meet point is equidistant to more than three points. Such a situation would be for example an intersection like the one in Figure 1.

A key feature that makes the GVG so useful for mobile robot navigation is that it can be constructed incrementally by using only sensor data and line of sight information.

Some meet points act as nodes on the graph and information about these nodes are required for the graph traversal. However, edges terminating in boundary points do not contain important information for the robot. The exploration of such paths leading to "dead-ends" does not result in any additional knowledge about the world. Using the sensor data, these edges are quite easily identified and removed from the GVG (see Figure 2).

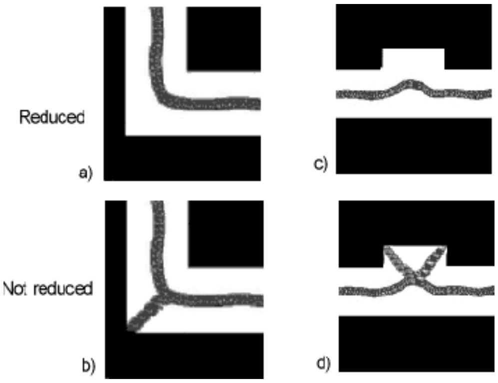


Fig. 2. *Reduction of the GVG.*

Let us see what kind of landmarks we use. Our robot is navigating in an indoor environment made by and for human beings. It means that we will find mainly vertical planes and a lot of vertical structures : doors, windows, posters on walls... That is what justifies our choice of planar, quadrangular landmarks. Furthermore, planar landmarks are stable and easily recognizable under very different viewpoints. We will assume that the camera optical axis remains mostly on a horizontal plane. A second assumption will be that our landmarks are delimited by two mostly vertical edges. As a consequence, we will typically find some posters on walls, some doors, but not necessarily physical entities, as viewed from a human point of view.

As far as 2D images primitives are concerned, several approaches have been studied to define visual landmarks : some authors use only segments whereas others rely on interest points (see [2] for instance). We will see that our method uses both and combines their advantages : segments stability and reliability, interest points frequency and discrimination power.

An exact calibration procedure for the camera is not required in order to check whether a region in an image corresponds to a planar landmark in the 3D scene [2, 17] when enough data is provided. It means for instance that we keep the opportunity to use the camera zooming abilities in an active framework.

3 Learning the topological graph

The exploration task consists in going over every path in the environment, memorizing the path connections in an GVG and learning some visual landmarks at the proximity to every meet point. From such a point at least two paths begin (two for a single corner, three or four for an intersection); a counter is maintained to keep track of unexplored paths in the "world". The counter starts at zero and when it comes back to zero the world is said to have been fully explored. Hereafter the different steps of the exploration task are listed.

Meet point detection. When it goes down an unexplored corridor, the robot is controlled to be on the GVG (between the two closer obstacles, typically the walls of a corridor); a meet point is detected if at least three points appear to be closer from the robot than the other ones in the same point vector acquired by the ultrasonic belt. Using the method proposed by Howie Choset [6], we compute the center of the meet point (point equidistant to the obstacles) and move towards it incrementally. More details can be found in [13].

Post Honing process. Once the meet point has been reached, and the honing phase is terminated, the robot has to decide whether this structural node is known or not. It makes a 360 degrees pan rotation constantly sending requests to the landmark detection module to see if any posters or other planar landmarks can be detected in the current image. If such a landmark is detected, then the robot finds its viewpoint-invariant representation to identify it (or not) and computes a rough angle at which it was found in the image (figure 3). The way the vision module builds the landmark representation for one or several detected landmarks is described in section 4. Let us note that a simpler and equivalent procedure could be used with an omni-directional visual sensor. After a detection, the robot position is altered away from the central point to the left or right to capture the landmark(s) at different angles. This is done to increase the confidence level in saliency and planarity, as it will be seen in 4.3. The robot then returns to its original position and continues to complete its 360 degree scan.

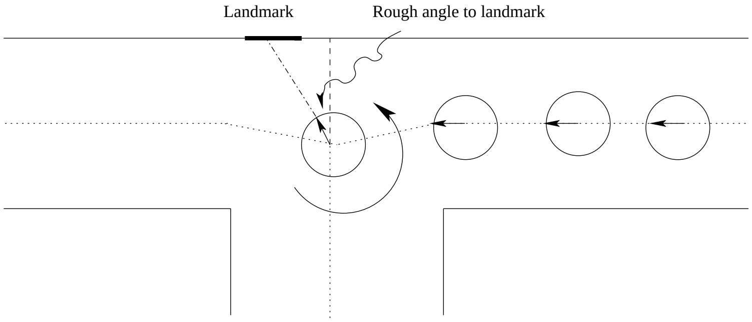


Fig. 3. *Meet point and visual landmark detection*

Add a new node in the model. On completion of a complete rotation if no landmarks were matched against the existing database of landmarks which were previously identified, the “best” landmark (highest saliency/stability) is selected and retained while the others are discarded. A new node ID is generated and the robot localizes itself in relation to the landmark : all departing paths from this node are defined with respect to its rough direction (see figure 4).

The paths are then set as unexplored for future exploration, except for the path from which the robot arrived at the current node. Connectivity information are updated : this arrival path for the current node points to the previous node and the departing path of the previous node points to the current node.

Recognition of a known node. If a landmark is found to be matched with a previously visited node, the robot localizes itself based on the landmark, and the paths are once again computed and matched against the previously identified paths. After this, explored paths are updated and the robot departs in one of the unexplored paths. If all paths from the node have been previously explored, the robot looks for the node at

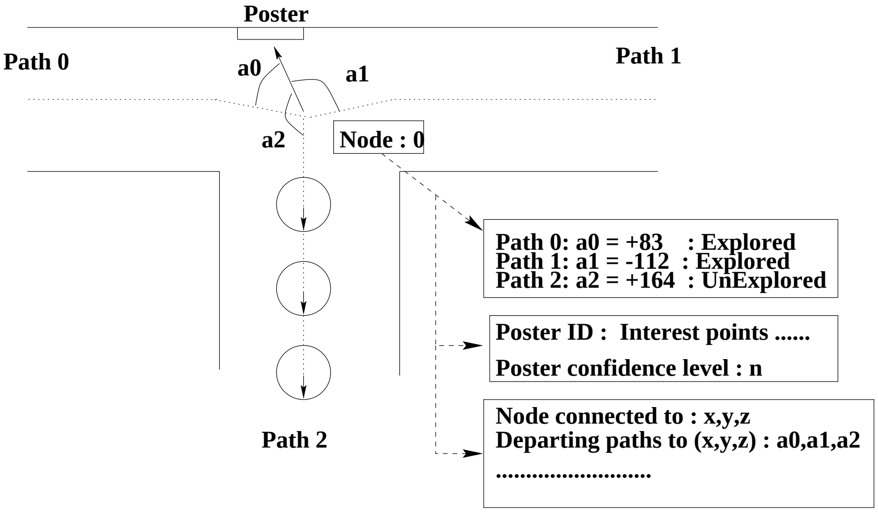


Fig. 4. *The paths from this node, are defined with respect to the landmark.*

the shortest distance from it with unexplored paths and moves in that direction.

4 Learning landmarks

To learn a model of a landmark on an autonomous way, we need on the one hand to set criteria and methods to detect this landmark and on the other hand to build a model that will be reusable for recognition.

4.1 Landmarks detection

The landmark detection mechanism is derived from the one explained in [10], inspired from studies about insect behaviours. The principle is illustrated in figure 5. We suppose that the structures present in our man-made environments will appear with mostly vertical edges in the image. This assumption allows us to accelerate the detection process. Note that we do not deal only with *posters*, but with *any quadrangle*, that may also be a door if this object is entirely visible.

The idea is to search the landmarks vertical edges on a 1D image resulting from an averaging operator on each grey levels column of the image. This signal is filtered and processed to roughly localize discontinuities; such a processing step is done with correlation with a step-like reference signal. Each of the selected discontinuity is potentially an edge of a landmark.

We use the rough information on the j-coordinate on each selected edge to get a full vertical segment in the 2D image. Indeed, the vertical lines corresponding to the selected j-coordinates are regularly sampled and step-like transitions around these sampled points are searched by correlation with step-like signals. After segmentation, we get the approximate positions (i, j) of the segments vertices. Figure 7 illustrates this approach.

Among all the selected vertically-oriented segments, indexed by their j-coordinates, we need to form couples (j_1, j_2) corresponding to potential landmarks. This matching procedure has been processed thanks to a discrete relaxation scheme. We define a priori probabilities for pairs $L_{kl} = (j_k, j_l)$ depending on the transition signs for both of the vertical profiles in segments j_k and j_l, on the distances between them and their relative vertical positions. We then define compatibilities between pairs of segments, so that, for example, we can introduce the fact that two potential landmarks may be related only by a relation of full inclusion or of no intersection, both in the i and j coordinate. Figure 6 shows the possible configurations for one or two quadrangles, whith a configuration example rejected by the relaxation scheme. All these constraints allow us to get a list of the most plausible potential landmarks.

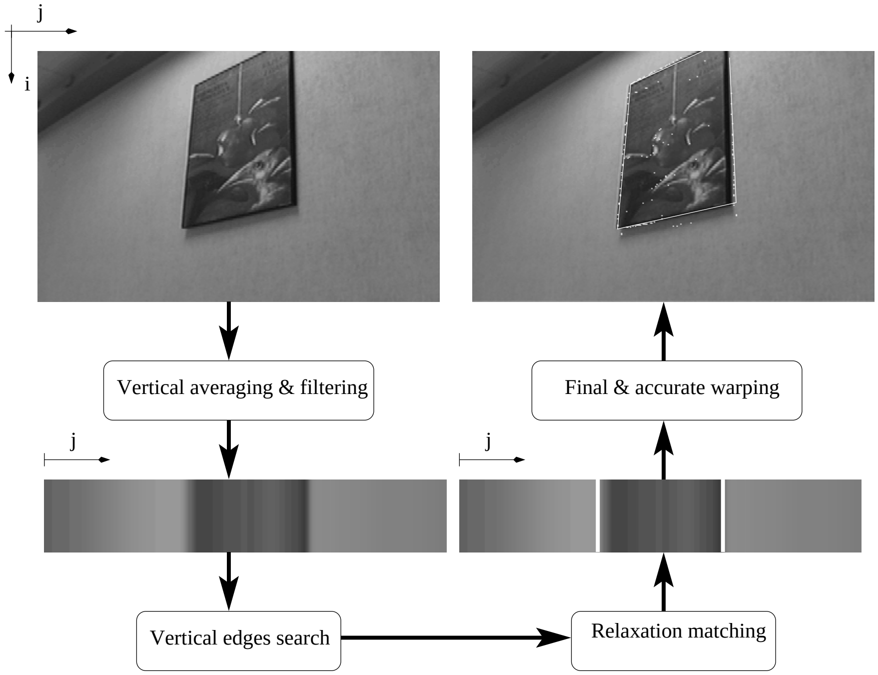


Fig. 5. Landmark detection.

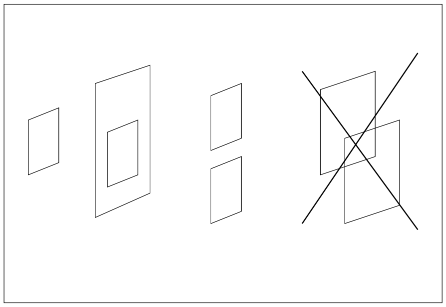


Fig. 6. Accepted configurations

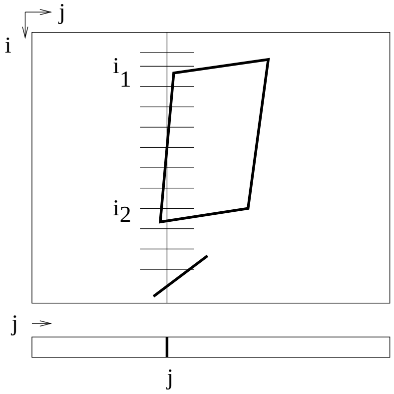


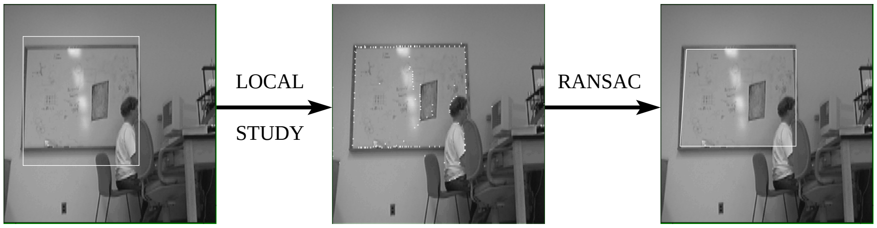
Fig. 7. From a 1D transition to a vertical segment

4.2 Landmarks accurate extraction

The principle, explained in figure 8, is the following one. The segments vertices give us four points and four estimated edges for the so-called landmark. Each of these four segments is regularly sampled and each sampled point will be considered as a *voting point*.

Then the procedure looks for the 1D grey level signal *along the normal direction to the segment* by computing correlation score with step-like reference signals. The maximum values of the correlation score give us sets of points more or less close to the landmark edges.

However, the 1D signal we deal with to get the straight line parameters may be very noisy : we must handle slight occlusions for instance. The *RANSAC* procedure is a voting scheme among all the correlation maxima we found before : this procedure allows us not to take into account the too noisy points for the computation of the straight line parameters : these parameters are computed only with the points that have "validated" the majoritary votes. In our implementation, 51% of the valid votes are necessary for the selection of given line parameters. "Valid" means that the vote is reliable and consistent : for instance, if the correlation coefficient is not sufficient, we do not take the vote into account.



Initial estimation **Voting points** **Selected straight lines**

Fig. 8. Landmarks accurate extraction

Figure 8 shows an example of extraction where some of the advantages of this method appear : for instance, the voting scheme allows to get the straight lines in spite of the board slight occlusion.

4.3 Planarity and saliency tests. Representation. Recognition process.

Now let us see how to select the best landmarks, how to represent them once they are extracted, how to use this representation to recognize them later and to update the information we know about them. Figure 10 sums up all this process on a simplified way.

In order to handle the perspective distorsion problem, the detected set of segments, a quadrangle, is first rectified : we compute an homography H between this quadrangle and a square with a given size (75×75 for instance). We will call this landmark representation, an "icon". H can be represented by a 3×3 matrix with 8 degrees of freedom, as relationships are written up to a scale. The computation of H is simply given by the 8 relationships between the icon vertices and the quadrangle ones, that can be written in a linear system. The resulting objects are shown in figure 9 both in real and synthetic images.

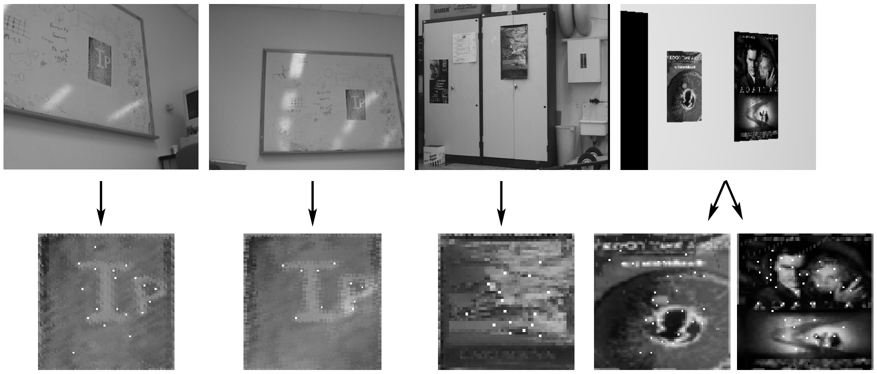


Fig. 9. Intrinsic representations for landmarks

We apply two kinds of saliency tests : the first one consists in computing a global covariance on the icon we found; the second one consists in finding representative corners with a Harris detector. If the global covariance or the number of found corners are inferior to a threshold (determined on experimental data), the landmark is considered as insufficiently salient and removed.

The set of detected corners will constitute the landmark model. Let us describe the recognition procedure between a reference landmark I_r and a current one I_n. The similarity between the sets of m_n and m_r corners P_n^k and P_r^l, $1 \leq k \leq m_n$, $1 \leq l \leq m_r$ resulting from I_n an I_r is computed thanks to a Hausdorff partial distance d defined according to :

$$d_1(n,r) = K_k^{th}(min_l||P_n^k, P_r^l||) \text{ and } d(n,r) = max(d_1(n,r), d_1(r,n))$$

If d is inferior to a given threshold, the poster n is "identified" as corresponding to the reference model r.

One can see that if d has a low value, it does not necessarily mean that landmark r is lying on a planar surface. Indeed, if we detect and recognize a landmark from small viewpoint differences with the reference view, we will surely have low values for d, even if it is not planar. To overcome this problem, we test at the end of the learning phase whether a given candidate landmark has been seen during the learning phase from sufficiently different viewpoints. This implies two things : first, a way to quantify the viewpoint changes; second, a method to ensure that the landmark will be seen at least from two different viewpoints.

To quantify the viewpoint changes between two views n_1 and n_2, we define the *viewpoint change measure* quantity $vcm(n_1, n_2)$:

$$vcm(n_1, n_2) = ||\hat{H}_{n_1 n_2} - I||$$

$\hat{H}$ is the homography between the two quadrangles in views n_1 and n_2 such that $\hat{H}_{33} = 1$. Then, we define a global *planarity confidence measure* over all the views $n_1, n_2 ... n_m$ corresponding to the landmark r by :

$$pcm(r) = \max_{k,l} vcm(k,l)$$

We use a threshold on the pcm value to reject landmarks on which planarity information, and consequently stability information too, are not sufficient.

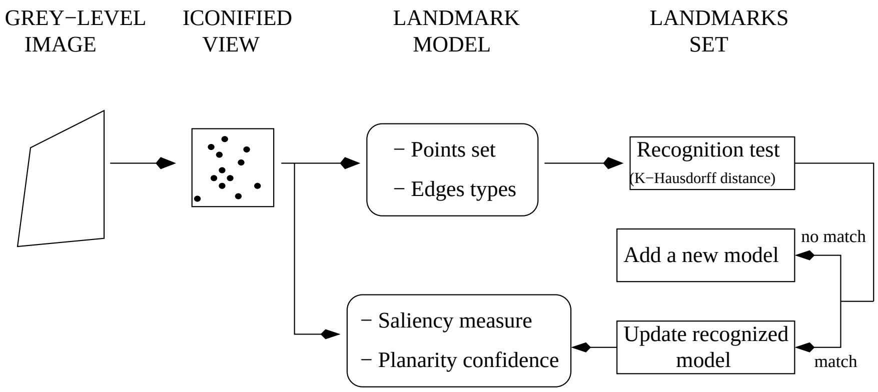


Fig. 10. Recognition and modelization

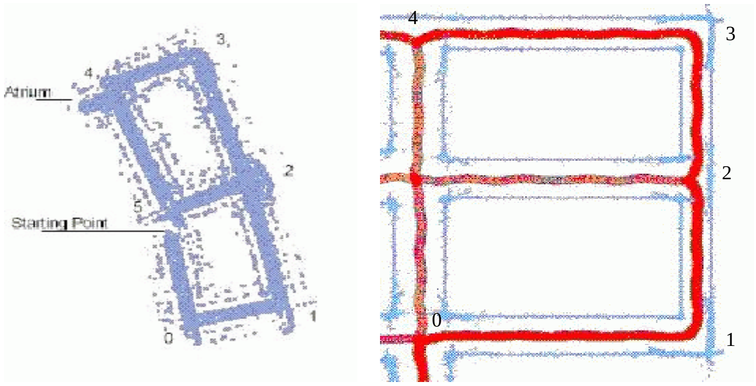
5 Experimental results

For the preliminary experiment in the Beckman Institute, only sonars were used; it has been shown that the GVG was well dedicated for mapping, path planning and exploration in one method. But, without vision, we have adopted a global localization technique based on odometer readings; we needed to minimize the number of wheel spins, and even with such a limitation, the node recognition procedure failed very often. This environment is shown in figure 11.

On the figure 12, experimental results for a complex building environment are presented. The environment is not a regular corridor network; only a partial exploration has been done (see the map on figure 12(a) : a long corridor with two posters, three doors, a corner and two entrances in a hallway). The robot is a Nomadic XR4000, equipped with a SICK laser range finder, two belts of ultrasonic sensors and a stereo rig mounted on a pan and tilt platform; this robot can be considered as an holonomic robot. By now, only the lower belt is used for the construction of the GVG model and images are acquired only from one static camera: the camera orientation is controlled using the robot rotation, without taking advantage of the pan degree of freedom.

Although only sonars are used for the GVG incremental construction, on figure 12(b), we display both sonar data (points) and laser segments with the robot trajectory. The robot finds two nodes in front of the two large entrances (even for the right entrance, only one node has been built, in spite of the corner proximity); from these nodes, the two posters are found and then, are associated to these nodes. But for the moment, the vision module is not activated during the corridor crossing, so that the three doors are not discovered.

The figure 12(b) shows clearly how noisy are the odometer measurements on the XR4000 robot. Our method relies on the odometry only when the robot X-axis is aligned with a poster axis, so that the path directions from a meet point can be expressed relatively to a landmark.



Results from real exploration **Results from simulation using accurate data readings**

Fig. 11. Experiments in Beckman Institute

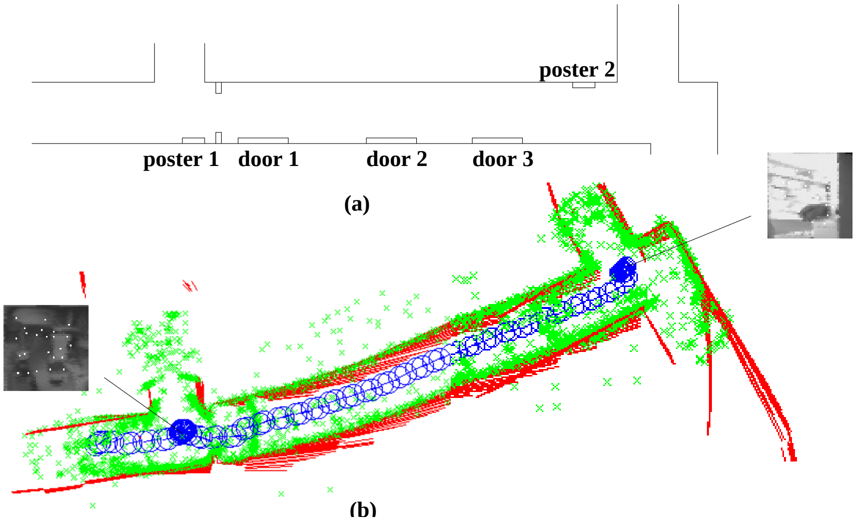


Fig. 12. *Experimental results : (a) the corridor map, (b) the exploration trajectory and the sensory data.*

6 Discussions and Future works

This paper has presented the integration of several topological based representations required for the navigation of a mobile robot in an office environment.

Our objective is to take advantage both from the Generalized Voronoi Graph model, suitable to represent a network of corridors, and from a landmark-based topological map which has been proposed to get rid of the classical problems which occur with an explicit self-localization with respect to an absolute reference frame. We avoid the use of traditionnal artificial visual landmarks by using, when available, some salient quadrangles in the scene.

In the current experiment, the environment is more complex than simple sets of orthogonal corridors, so that the GVG approach is useful, whereas vision is mandatory to guarantee a good recognition of the nodes.

We are currently trying to get some more significative experimental results for this complex environment; however, when dealing with a mixture of corridors, rooms and open spaces, the geometrical characterization of the meet points using sonars is difficult and visual landmarks must be considered to support local reference frames.

In our future works, we intend to improve the exploration task by the use of a Laser Range Finder which gives better measurements than the ultrasonic sensors, and to add an uncertainty representation to our model using Hidden Markov Models [7].

References

1. H. Bulata and M. Devy. Incremental construction of a landmark-based and topological model of indoor environments by a mobile robot. In *Proc. 1996 IEEE International Conference on Robotics and Automation (ICRA'96)*, Minneapolis, USA, April 1996.
2. C.I. Colios and P.E. Trahanias. Landmark identification based on projective and permutation invariant vectors. In *ICPR'2000*, pages 128–131, September 2000.
3. G. Dudek and P. MacKenzie. Model-Based Map Construction for Robot Localization. In *Vision Interface*, 1993.
4. H. Durrant-White. *Integration, Coordination and Control of multisensor robot systems.* Kluwer Academy Pub., 1988.
5. J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omni-directionnalcamera. *IEEE Trans. on Robotics and Automation*, 16(6), December 2000.
6. Keiji Nagatani Howie Choset and Alfred Rizzi. Sensor based planning: Using a honing strategy and local map method to implement the generalized voronoi graph. In *SPIE Mobile Robotics, Pittsburgh, PA*, 1997.
7. S. Koenig and R. Simmons. Unsupervised learning of probabilistic models for robot navigation. In *IEEE Int. Conf. on Robotics and Automation (ICRA '96)*, pages 2301–2308, Minneapolis (USA), 1996.
8. D. Kortenkampf and T. weymouth. Topological mapping for mobile robots using a combination of sonar and vision sensing. In *AAAI*, July 1994.
9. B. Kuipers and Y. T. Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Journal of Robotics and Autonomous Systems*, 8(1-2), 1991.
10. D. Lambrinos, R. Möller, T. Labhart, R. Pfeifer, and R. Wehner. A mobile robot employing insect strategies for navigation. *NeuroComputing*, 1999.
11. M.Devy and C.Parra. Topological modelling for outdoor mobile robots. In *Proc. 7th International Symposium on Intelligent Robotic Systems (SIRS'99), Rapport LAAS No99144*, Coimbra, Portugal, July 1999.
12. K. Nagatani, H. Choset, and S.Thrun. Towards exact localization without explicit localization with the generalized voronoï graph. In *IEEE International Conference on Robotics and Automation(ICRA'98),Louvain (Belgique)*, May 1998.
13. P. Ranganathan, J.B. Hayet, M. Devy, S. Hutchinson, and F. Lerasle. Incremental construction of a topological model for indoor environments using visual and sonar data. Submitted to IEEE/RSJ Int. Conf. on Intelligent Robots and Systems IROS'2001.
14. J. Santos-Victor, R. Vassallo, and H.J. Schneebeli. Topological maps for visual navigation. In *1st International Conference on Computer Vision Systems*, Las Palmas, Spain, January 1999.
15. R.C. Smith, M. Self, and P. Cheeseman. Estimating Uncertain Spatial Relationships in Robotics. *Autonomous Robot Vehicules*, pages 167–193, 1990.
16. C. Taylor and D. Kriegman. Vision-Based Motion Planning and Exploration Algorithms for Mobile Robots. *Algorithmic Foundations of Robotics*, 1995.
17. B. Triggs. Autocalibration from planar scenes. In *ECCV'98*, 1998.
18. D. Van Zwynsvoorde, T. Simeon, and R. Alami. Incremental topological modeling using local voronoï-like graphs. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems IROS'2000*, Takamatsu, Japan, 2000.

This article was processed using
the TEX macro package with SIRS2001 style